

Unbiased Equation-Error based Algorithms for Efficient System Identification using Noisy Measurements

H. C. So*, Y. T. Chan[†], K. C. Ho[‡] and Frankie K. W. Chan*

*Department of Electronic Engineering, City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong

[†]Department of Electrical and Computer Engineering, Royal Military College of Canada
Kingston, Ontario, Canada K7K 7B4

[‡]Department of Electrical and Computer Engineering, University of Missouri-Columbia
Columbia, MO 65211, USA

September 27, 2006

Index terms: *Infinite impulse response filtering, equation error approach, weighted least squares, unit-norm constraint, monic constraint*

Abstract

Based on the equation-error approach, two constrained weighted least squares algorithms are developed for unbiased infinite impulse response system identification. Both white input and output noise are present, and the ratio of the noise powers is known. Through a weighting matrix, the first algorithm uses a generalized unit-norm constraint which is a generalization of the Koopmans-Levin method. The second method employs a monic constraint which in fact is a relaxation algorithm for maximum likelihood estimation in Gaussian noise. Algorithm modifications for the input-noise-only or output-noise-only cases are also given. Via computer simulations, the effectiveness of the proposed estimators is demonstrated by contrasting with conventional benchmarks in different signal-to-noise ratio and data length conditions.

Corresponding Author

H. C. So

Department of Electronic Engineering

City University of Hong Kong

Tat Chee Avenue, Kowloon, Hong Kong

Tel. : (852) 2788 7780

Fax: (852) 2788 7791

email: hcso@ee.cityu.edu.hk

1 Introduction

The problem of identifying linear systems from their input and output measurements has received significant attention because of its important applications in signal processing, communications and control [1]-[4]. In this study, we consider infinite impulse response (IIR) system identification in the presence of white input and/or output measurement noise, which is also known as the errors-in-variables (EIV) problem [4]. Standard techniques for system identification include joint output (JO) approach [5], Frisch scheme [6]-[7], structured total least squares (STLS) [8]-[10] and Koopmans-Levin (KL) [11]-[14] approach. One major advantage of the JO and Frisch methods are that they can estimate the unknown input and output noise powers as well. While the STLS and KL methods assume that the ratio of the noise powers is available. In this paper, we will formulate the EIV problem using the equation-error (EE) [15]-[16] approach which has the major advantage of global convergence, although compensation of the parameter bias is required to achieve unbiased estimation.

In the presence of white output noise only, Regalia [16] has proposed to minimize the EE cost function subject to a unit-norm constraint. Recently, it has been shown [17]-[18] that the constrained EE optimization can be formulated in a mixed least squares (LS)-total least squares framework. Furthermore, the KL scheme, which can also be cast in the EE framework, provides unbiased estimates by finding the eigenvector corresponding to the minimum eigenvalue of the sample covariance matrix [12], or the right singular vector corresponding to the smallest singular value of the data matrix [13]-[14]. However, all the above EE schemes only use standard LS optimization and they generally cannot produce optimum estimation performance, that is, minimum achievable mean square errors of the parameter estimates. In this paper, our major contribution is to utilize the technique of weighted least squares (WLS), together with appropriate constraints, to boost the estimation accuracy of the EE approach for system identification.

The rest of the paper is organized as follows. In Section 2, we first consider solving the EIV problem with the use of LS, assuming that the noise power ratio is available. An improvement to the LS solution is then devised by minimizing a WLS cost function subject to a generalized unit-norm constraint, where the weighting matrix is determined iteratively. The second algorithm we develop also employs WLS but with the monic constraint. Simulation results are presented in Section 3 to evaluate the proposed algorithms by comparing with different benchmarks, namely, the LS, KL, output-error (OE) [3], prediction-error (PE) [2] and STLS methods [9] as well as the asymptotic Cramér-Rao lower bound (CRLB) for infinite measurements [20]. It is shown that the mean square errors of the proposed WLS algorithms attain the CRLB for sufficiently high signal-to-noise ratios (SNRs) and/or large data lengths. Finally, concluding remarks are drawn in Section 4.

2 Algorithm Development

In the following, the problem of IIR system identification in the presence of input and output noise based on the EE formulation is first reviewed. A unit-norm LS solution via minimization of the EE-based LS cost function subject to a generalized unit-norm constraint is obtained, which is identical to the KL method. This approach is then improved by using WLS with an iterative procedure. The second approach also employs WLS but with the monic constraint. Finally, we present the required modifications of the proposed unit-norm WLS algorithm for the output-noise-only systems and the results can be readily applied to the input-noise-only systems as well as to the monic WLS algorithm.

2.1 Equation-Error Formulation

The unknown IIR system is stable and causal and has the form

$$H(z) = \frac{B(z)}{A(z)} \quad (1)$$

where

$$A(z) = a_0 + \sum_{l=1}^L a_l z^{-l}$$

$$B(z) = - \sum_{l=0}^L b_l z^{-l}$$

with $a_0 = 1$. Let the orders of the denominator and numerator polynomials be known, and without loss of generality, have identical values of L . The observed system input and output are

$$x_k = s_k + m_k \quad (2)$$

$$r_k = d_k + n_k, \quad k = 0, 1, \dots, N-1$$

where s_k and d_k denote the noise-free input and output, respectively, and m_k and n_k represent the input and output measurement noise which are independent of s_k . It is assumed that s_k is a stochastic process while m_k and n_k are uncorrelated zero-mean white processes with unknown variances σ_m^2 and σ_n^2 , respectively, but the ratio of the noise powers, say, $\alpha^2 = \sigma_n^2/\sigma_m^2$, is available [18], [13]-[14]. Given the N samples of x_k and r_k , the task is to find a_l , $l = 1, 2, \dots, L$, and b_l , $l = 0, 1, \dots, L$.

In the EE formulation, the error function e_k has the form

$$e_k = \sum_{l=0}^L \tilde{a}_l r_{k-l} + \sum_{l=0}^L \tilde{b}_l x_{k-l} \quad (3)$$

where $\{\tilde{a}_l\}$ and $\{\tilde{b}_l\}$ are optimization variables of $\{a_l\}$ and $\{b_l\}$ up to a scalar because we have not assigned $\tilde{a}_0 = 1$.

2.2 Weighted Least Squares with Generalized Unit-Norm Constraint

The corresponding LS cost function is then

$$J(\tilde{\boldsymbol{\rho}}) = \mathbf{e}^T \mathbf{e} = \tilde{\boldsymbol{\rho}}^T \mathbf{Y}^T \mathbf{Y} \tilde{\boldsymbol{\rho}} \quad (4)$$

where

$$\mathbf{e} = [e_{N-1}, e_{N-2}, \dots, e_L]^T$$

$$\mathbf{Y} = \begin{bmatrix} r_{N-1} & r_{N-2} & \cdots & r_{N-L-1} & x_{N-1} & x_{N-2} & \cdots & x_{N-L-1} \\ r_{N-2} & r_{N-3} & \cdots & r_{N-L-2} & x_{N-2} & x_{N-3} & \cdots & x_{N-L-2} \\ \vdots & \vdots \\ r_L & r_{L-1} & \cdots & r_0 & x_L & x_{L-1} & \cdots & x_0 \end{bmatrix}$$

$$\tilde{\boldsymbol{\rho}} = [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_L, \tilde{b}_0, \tilde{b}_1, \dots, \tilde{b}_L]^T$$

and T denotes the transpose operation. It is easy to show that the expected value of $J(\tilde{\boldsymbol{\rho}})$ is

$$\begin{aligned} E\{J(\tilde{\boldsymbol{\rho}})\} &= E\{(\mathbf{S}\tilde{\boldsymbol{\rho}})^T \mathbf{S}\tilde{\boldsymbol{\rho}}\} + \tilde{\boldsymbol{\rho}}^T E\{\mathbf{Q}^T \mathbf{Q}\} \tilde{\boldsymbol{\rho}} \\ &= E\{\tilde{\boldsymbol{\rho}}^T \mathbf{S}^T \mathbf{S} \tilde{\boldsymbol{\rho}}\} + (N-L)\tilde{\boldsymbol{\rho}}^T \boldsymbol{\Sigma} \tilde{\boldsymbol{\rho}} \sigma_m^2 \end{aligned} \quad (5)$$

where

$$\boldsymbol{\Sigma} = \text{diag}(\underbrace{\alpha^2, \dots, \alpha^2}_{L+1}, \underbrace{1, \dots, 1}_{L+1})$$

$$\mathbf{S} = \begin{bmatrix} d_{N-1} & d_{N-2} & \cdots & d_{N-L-1} & s_{N-1} & s_{N-2} & \cdots & s_{N-L-1} \\ d_{N-2} & d_{N-3} & \cdots & d_{N-L-2} & s_{N-2} & s_{N-3} & \cdots & s_{N-L-2} \\ \vdots & \vdots \\ d_L & d_{L-1} & \cdots & d_0 & s_L & s_{L-1} & \cdots & s_0 \end{bmatrix}$$

and

$$\mathbf{Q} = \begin{bmatrix} n_{N-1} & n_{N-2} & \cdots & n_{N-L-1} & m_{N-1} & m_{N-2} & \cdots & m_{N-L-1} \\ n_{N-2} & n_{N-3} & \cdots & n_{N-L-2} & m_{N-2} & m_{N-3} & \cdots & m_{N-L-2} \\ \vdots & \vdots \\ n_L & n_{L-1} & \cdots & n_0 & m_L & m_{L-1} & \cdots & m_0 \end{bmatrix}$$

The \mathbf{S} and \mathbf{Q} represent the signal and noise components of \mathbf{Y} , respectively, such that $\mathbf{Y} = \mathbf{S} + \mathbf{Q}$.

We observe that when $\{\tilde{a}_l\}$ and $\{\tilde{b}_l\}$ equal the true system parameters, then $\mathbf{S}\tilde{\boldsymbol{\rho}}$ is a zero vector.

However, the minimum of $E\{J(\tilde{\boldsymbol{\rho}})\}$ will not correspond to the desired $\{a_l\}$ and $\{b_l\}$ in the presence of

input and/or output noise because the second term of (5) also contains the unknown parameters to be estimated. Extending the idea of the unit-norm constraint approach for output-noise-only systems [16] to noisy input-output systems, unbiased IIR system identification can be achieved via minimizing $E\{J(\tilde{\boldsymbol{\rho}})\}$ subject to

$$\tilde{\boldsymbol{\rho}}^T \boldsymbol{\Sigma} \tilde{\boldsymbol{\rho}} = 1 \quad (6)$$

This is a generalized unit-norm constraint and we refer this scheme to as the unit-norm LS method. Using the technique of Lagrange multipliers, the constrained optimization problem can be solved by computing the generalized eigenvector corresponding to the smallest generalized eigenvalue of the pair $(\mathbf{Y}^T \mathbf{Y}, \boldsymbol{\Sigma})$. Comparing with [13]-[14], it is easy to see that the unit-norm LS method is in fact identical to the KL solution computed from the data matrix, although we use eigenvalue decomposition while the latter is realized by singular value decomposition (SVD).

Since it is well known [19] that the KL approach is inferior to the maximum likelihood (ML) method, the unit-norm LS method will also produce suboptimum parameter estimates as well. Nevertheless, the formulation of the unit-norm LS approach leads directly to an improvement through adding a symmetric positive definite matrix, say, \mathbf{W} , to the LS cost function. An optimal choice of \mathbf{W} is the Markov estimate [3]:

$$\mathbf{W} = \sigma_m^2 (E\{\boldsymbol{\epsilon} \cdot \boldsymbol{\epsilon}^T\})^{-1} \quad (7)$$

where

$$\boldsymbol{\epsilon} = [\epsilon_{N-1}, \epsilon_{N-2}, \dots, \epsilon_L]^T$$

and

$$\epsilon_k = \sum_{l=0}^L \tilde{a}_l n_{k-l} + \sum_{l=0}^L \tilde{b}_l m_{k-l}$$

being the residual error. The inverse of \mathbf{W} is shown to be

$$\mathbf{W}^{-1} = \mathbf{G}^T \boldsymbol{\Sigma}_1 \mathbf{G} \quad (8)$$

where

$$\mathbf{G}^T = \begin{bmatrix} \tilde{a}_L & \tilde{a}_{L-1} & \cdots & \tilde{a}_0 & 0 & 0 & \cdots & 0 & \tilde{b}_L & \tilde{b}_{L-1} & \cdots & \tilde{b}_0 & 0 & 0 & \cdots & 0 \\ 0 & \tilde{a}_L & \cdots & \tilde{a}_1 & \tilde{a}_0 & 0 & \cdots & 0 & 0 & \tilde{b}_L & \cdots & \tilde{b}_1 & \tilde{b}_0 & 0 & \cdots & 0 \\ \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \tilde{a}_1 & \tilde{a}_0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \tilde{b}_1 & \tilde{b}_0 \end{bmatrix}$$

is of dimension $(N - L) \times N$ and $\boldsymbol{\Sigma}_1 = \text{diag}(\underbrace{\alpha^2, \dots, \alpha^2}_N, \underbrace{1, \dots, 1}_N)$. With weighting matrix, the cost function is now:

$$J_{\text{unit-norm}}(\tilde{\boldsymbol{\rho}}) = \mathbf{e}^T \mathbf{W} \mathbf{e} = \tilde{\boldsymbol{\rho}}^T \mathbf{Y}^T \mathbf{W} \mathbf{Y} \tilde{\boldsymbol{\rho}} \quad (9)$$

From (4), (5) and (9), it is easy to show that for WLS minimization, the corresponding constraint will become:

$$\tilde{\boldsymbol{\rho}}^T \boldsymbol{\Upsilon} \tilde{\boldsymbol{\rho}} = 1 \quad (10)$$

where $\boldsymbol{\Upsilon} = E\{\mathbf{Q}^T \mathbf{W} \mathbf{Q}\} / \sigma_m^2$ is a matrix of dimension $(2L + 2) \times (2L + 2)$ and has the form:

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \alpha^2 D_0 & \alpha^2 D_1 & \cdots & \alpha^2 D_{L-1} & \alpha^2 D_L & 0 & 0 & \cdots & 0 & 0 \\ \alpha^2 D_1 & \alpha^2 D_2 & \cdots & \alpha^2 D_L & \alpha^2 D_{L-1} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots \\ \alpha^2 D_L & \alpha^2 D_{L-1} & \cdots & \alpha^2 D_1 & \alpha^2 D_0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & D_0 & D_1 & \cdots & D_{L-1} & D_L \\ 0 & 0 & \cdots & 0 & 0 & D_1 & D_2 & \cdots & D_L & D_{L-1} \\ \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & D_L & D_{L-1} & \cdots & D_1 & D_0 \end{bmatrix} \quad (11)$$

where $D_j = \sum_{i=1}^{N-L-j} [\mathbf{W}]_{i,i+j}$, $j = 0, 1, \dots, L$, with $[\mathbf{W}]_{i,j}$ representing the (i, j) -entry of \mathbf{W} . It is observed that (4) and (6) are only a special case of (9) and (10) when \mathbf{W} is the identity matrix. Since \mathbf{W} and $\boldsymbol{\Upsilon}$ are unknown, we propose to estimate $\boldsymbol{\rho}$ in an iterative manner as follows:

- (i) Find an initial estimate of the system parameter vector, denoted by $\hat{\boldsymbol{\rho}}$, through computing the generalized eigenvector corresponding to the minimum generalized eigenvalue of the pair $(\mathbf{Y}^T \mathbf{Y}, \boldsymbol{\Sigma})$, which is the unbiased unit-norm LS solution.
- (ii) Use the estimated $\hat{\boldsymbol{\rho}}$ to construct $\hat{\mathbf{W}}$ as well as $\hat{\boldsymbol{\Upsilon}}$, which represent the estimates of \mathbf{W} and $\boldsymbol{\Upsilon}$, respectively, based on (8) and (11).
- (iii) Compute the generalized eigenvector corresponding to the minimum generalized eigenvalue of the pair $(\mathbf{Y}^T \hat{\mathbf{W}} \mathbf{Y}, \hat{\boldsymbol{\Upsilon}})$, which is the updated parameter estimate $\hat{\boldsymbol{\rho}}$.
- (iv) Repeat Steps (ii) and (iii) for a few iterations until parameter convergence. Then divide the eigenvector by \hat{a}_0 to get the IIR coefficient estimates.

We refer this algorithm to as the unit-norm WLS estimator. Although there is no guarantee of convergence for the above relaxation algorithm [21], our simulation studies in Section 3 demonstrate that the unit-norm WLS estimator is able to achieve global convergence with optimum estimates for sufficiently large SNRs and/or data lengths.

2.3 Weighted Least Squares with Monic Constraint

In the literature, another commonly used constraint for linear system identification is the monic constraint [15]-[16], [22]-[23]. In the following, the WLS algorithm with monic constraint of $\tilde{a}_0 = 1$ for

IIR estimation is developed based on (9). First, we partition $\tilde{\boldsymbol{\rho}}$ and \mathbf{Y} as

$$\tilde{\boldsymbol{\rho}} = [1 \quad \tilde{\boldsymbol{\varrho}}^T]^T \quad (12)$$

where

$$\tilde{\boldsymbol{\varrho}} = [\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_L, \dots, \tilde{b}_0, \tilde{b}_1, \dots, \tilde{b}_L]^T$$

and

$$\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2] \quad (13)$$

where

$$\mathbf{Y}_1 = \begin{bmatrix} r_{N-1} \\ r_{N-2} \\ \vdots \\ r_L \end{bmatrix}$$

and

$$\mathbf{Y}_2 = \begin{bmatrix} r_{N-2} & r_{N-3} & \cdots & r_{N-L-1} & x_{N-1} & x_{N-2} & \cdots & x_{N-L-1} \\ r_{N-3} & r_{N-4} & \cdots & r_{N-L-2} & x_{N-2} & x_{N-3} & \cdots & x_{N-L-2} \\ \vdots & \vdots \\ r_{L-1} & r_{L-2} & \cdots & r_0 & x_L & x_{L-1} & \cdots & x_0 \end{bmatrix}$$

Substituting (12) and (13) into (9) yields the monic WLS cost function

$$J_{\text{monic}}(\tilde{\boldsymbol{\rho}}) = (\mathbf{Y}_1 + \mathbf{Y}_2 \tilde{\boldsymbol{\varrho}})^T \mathbf{W} (\mathbf{Y}_1 + \mathbf{Y}_2 \tilde{\boldsymbol{\varrho}}) \quad (14)$$

where $\tilde{a}_0 = 1$ in \mathbf{W} as well. It can be shown that the minimum of (14) corresponds to the ML estimate of $\boldsymbol{\varrho} = [a_1, a_2, \dots, a_L, \dots, b_0, b_1, \dots, b_L]^T$ in white Gaussian noise. However, ML estimation will involve extensive computation as \mathbf{W} is a function of $\tilde{\boldsymbol{\varrho}}$ as well. In our study, we substitute $\hat{\mathbf{W}}$ for \mathbf{W} so that $J_{\text{monic}}(\tilde{\boldsymbol{\varrho}})$ is now a quadratic function in $\tilde{\boldsymbol{\varrho}}$. Differentiating the approximate version of $J_{\text{monic}}(\tilde{\boldsymbol{\varrho}})$ with respect to $\tilde{\boldsymbol{\varrho}}$ and then setting the resultant expression to zero, we obtain our monic WLS solution:

$$\hat{\boldsymbol{\varrho}} = -(\mathbf{Y}_2^T \hat{\mathbf{W}} \mathbf{Y}_2)^{-1} \mathbf{Y}_2^T \hat{\mathbf{W}} \mathbf{Y}_1 \quad (15)$$

It is expected that the WLS estimator with the monic constraint will attain nearly optimum performance when $\hat{\mathbf{W}} \rightarrow \mathbf{W}$. As in the unit-norm WLS algorithm, $\hat{\mathbf{W}}$ is not available at the beginning and we propose the following iterative estimation procedure:

- (i) Use the unbiased unit-norm LS solution as the initial estimate of the system parameter vector $\hat{\boldsymbol{\rho}}$. Normalize the first element of $\hat{\boldsymbol{\rho}}$ to be unity. Note that it is possible to substitute $\hat{\mathbf{W}}$ with the identity matrix in (15) for parameter initialization [21]. However, the corresponding solution is biased because \mathbf{Y}_1 is correlated with \mathbf{Y}_2 [3], which may increase the possibility of local convergence and subsequently results in a smaller SNR operation range than that of the unit-norm LS solution.
- (ii) Use the estimated $\hat{\boldsymbol{\rho}}$ to construct $\hat{\mathbf{W}}$ based on (8).
- (iii) Compute an improved $\hat{\boldsymbol{\rho}}$ using (15).
- (iv) Repeat Steps (ii) and (iii) for a few iterations until parameter convergence. It is worthy to mention that the iterative reweighting procedure can eliminate the estimator bias upon convergence, even though the initial parameter estimates may be biased.

We refer this algorithm to as the monic WLS algorithm. Again, we are not aware [21] of the existence of a proof of global convergence for the above relaxation algorithm but its excellent estimation performance has been demonstrated via computer simulations in Section 3.

2.4 Modifications for Output-Noise-Only Systems

In this Section, we modify the unit-norm WLS algorithm for the output-noise-only case and the results can be straightforwardly applied to the input-noise-only case as well as to the monic WLS algorithm. In fact, system identification with only output noise is also frequently addressed in the literature [1], [16] and commercial available algorithms [24] for this problem including the OE and PE methods, where the former is optimum only for white Gaussian distributed data while the optimality of the latter holds for general Gaussian processes [3]. As a result, the following development also allows us to compare the proposed WLS approach with the OE and PE algorithms.

The weighting matrix \mathbf{W} is now

$$\mathbf{W}^{-1} = \mathbf{G}^T \mathbf{G} \tag{16}$$

where

$$\mathbf{G}^T = \begin{bmatrix} \tilde{a}_L & \tilde{a}_{L-1} & \cdots \tilde{a}_0 & 0 & 0 & \cdots & 0 \\ 0 & \tilde{a}_L & \cdots \tilde{a}_1 & \tilde{a}_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \tilde{a}_1 & \tilde{a}_0 \end{bmatrix}$$

which does not involve the unknown parameter σ_n^2 . Following Section 2.2, the generalized unit-norm

constraint becomes

$$[\tilde{\mathbf{a}}^T \ \tilde{\mathbf{b}}^T] \begin{bmatrix} \mathbf{\Upsilon}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{b}} \end{bmatrix} = 1 \quad (17)$$

where $\tilde{\mathbf{a}} = [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_L]^T$, $\tilde{\mathbf{b}} = [\tilde{b}_0, \tilde{b}_1, \dots, \tilde{b}_L]^T$, $\mathbf{0}$ represents a zero matrix of dimension $(L+1) \times (L+1)$, and $\mathbf{\Upsilon}_{11}$ is also of dimension $(L+1) \times (L+1)$ and has the form

$$\mathbf{\Upsilon}_{11} = \begin{bmatrix} D_0 & D_1 & \cdots & D_{L-1} & D_L \\ D_1 & D_2 & \cdots & D_L & D_{L-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_L & D_{L-1} & \cdots & D_1 & D_0 \end{bmatrix} \quad (18)$$

Using the technique of Lagrange multipliers, the Lagrangian to be minimized is

$$\begin{aligned} \mathcal{L}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \lambda) &= [\tilde{\mathbf{a}}^T \ \tilde{\mathbf{b}}^T] \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{b}} \end{bmatrix} + \lambda \left(1 - [\tilde{\mathbf{a}}^T \ \tilde{\mathbf{b}}^T] \begin{bmatrix} \mathbf{\Upsilon}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{b}} \end{bmatrix} \right) \\ &= \tilde{\mathbf{a}}^T \mathbf{Z}_{11} \tilde{\mathbf{a}} + 2\tilde{\mathbf{a}}^T \mathbf{Z}_{12} \tilde{\mathbf{b}} + \tilde{\mathbf{b}}^T \mathbf{Z}_{22} \tilde{\mathbf{b}} + \lambda (1 - \tilde{\mathbf{a}}^T \mathbf{\Upsilon}_{11} \tilde{\mathbf{a}}) \end{aligned} \quad (19)$$

where \mathbf{Z}_{11} , \mathbf{Z}_{12} , \mathbf{Z}_{21} and \mathbf{Z}_{22} are partitioned matrices of $\mathbf{Y}^T \mathbf{W} \mathbf{Y}$ with $\mathbf{Z}_{12} = \mathbf{Z}_{21}^T$, and λ is the Lagrange multiplier. Differentiating $\mathcal{L}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \lambda)$ with respect to $\tilde{\mathbf{a}}$ and then setting the resultant expression to zero yields

$$\hat{\mathbf{Z}}_{11} \hat{\mathbf{a}} + \hat{\mathbf{Z}}_{12} \hat{\mathbf{b}} - \lambda \mathbf{\Upsilon}_{11} \hat{\mathbf{a}} = \mathbf{0} \quad (20)$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are the estimates of \mathbf{a} and \mathbf{b} , respectively. Similarly, we differentiate $\mathcal{L}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \lambda)$ with respect to $\tilde{\mathbf{b}}$ and set the resultant expression to zero to obtain

$$\hat{\mathbf{b}} = -\mathbf{Z}_{22}^{-1} \mathbf{Z}_{12}^T \hat{\mathbf{a}} \quad (21)$$

Substituting (21) into (20), we have

$$\left(\mathbf{Z}_{11} - \mathbf{Z}_{12} \mathbf{Z}_{22}^{-1} \mathbf{Z}_{12}^T \right) \hat{\mathbf{a}} = \lambda \mathbf{\Upsilon}_{11} \hat{\mathbf{a}} \quad (22)$$

That is, $\hat{\mathbf{a}}$ is determined as the generalized eigenvector corresponding to the minimum generalized eigenvalue of the pair $(\mathbf{Z}_{11} - \mathbf{Z}_{12} \mathbf{Z}_{22}^{-1} \mathbf{Z}_{12}^T, \mathbf{\Upsilon}_{11})$. Once we have found $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ is then obtained from (21). As a result, the relaxation algorithm of the unit-norm WLS estimator for output-noise-only systems is given as follows:

- (i) Find initial estimates of the system parameters by computing the generalized eigenvector corresponding to the minimum generalized eigenvalue of the pair $(\mathbf{Z}_{11} - \mathbf{Z}_{12} \mathbf{Z}_{22}^{-1} \mathbf{Z}_{12}^T, \mathbf{\Upsilon}_{11})$ with $\mathbf{\Upsilon}_{11}$ equal to the identity matrix, which are the unit-norm solution proposed in [16].
- (ii) Use the estimated $\hat{\boldsymbol{\rho}}$ to construct $\hat{\mathbf{W}}$ as well as $\hat{\mathbf{\Upsilon}}_{11}$ based on (16) and (18), respectively.

- (iii) Compute the generalized eigenvector corresponding to the minimum generalized eigenvalue of the pair $(\mathbf{Z}_{11} - \mathbf{Z}_{12}\mathbf{Z}_{22}^{-1}\mathbf{Z}_{12}^T, \hat{\mathbf{Y}}_{11})$.
- (iv) Repeat Steps (ii) and (iii) for a few iterations until parameter convergence.
- (v) Normalize $\hat{\mathbf{a}}$ with $\hat{a}_0 = 1$ and then compute $\hat{\mathbf{b}}$ using (21).

3 Numerical Examples

Computer simulations using MATLAB had been performed to evaluate the performance of the two proposed algorithms for estimating IIR system parameters. The noise-free input s_k was a sequence of zero-mean, independent identically distributed Gaussian random variables with unity power while m_k and n_k were also independent white Gaussian processes. The noisy input-output, and noisy output only cases were considered. In the former, we assigned $\alpha^2 = 20$ and comparison was made with the unit-norm LS, SVD-based KL [13]-[14] and STLS [9] methods as well as asymptotic CRLB for Gaussian input and white Gaussian disturbances [20]. While for the latter, we contrasted the developed estimators with several optimum output-noise-only system identification methods, namely, the OE, PE [2]-[3] using [24] as well as the STLS scheme. The unknown system had the form

$$H(z) = \frac{1}{1 - z^{-1} + 0.5z^{-2}}$$

which was second order with parameters $a_1 = -1$, $a_2 = 0.5$ and $b_0 = -1$. For the iterative procedures of the unit-norm WLS and monic WLS algorithms, we stopped after three iterations because it was found that the algorithms almost converged with this termination criterion. The standard performance measures of mean and mean square error (MSE) were used and all results were obtained from 500 independent runs.

Figures 1 to 3 show the MSEs for a_1 , a_2 and b_0 , respectively, of the unit-norm LS, unit-norm WLS, monic WLS, KL and STLS methods in estimating the noisy input-output system at output noise power ranging from -30 dB to 20 dB with $N = 50$. When the output noise power was smaller than -5 dB, the two WLS techniques had MSEs whose values were close to the CRLB, and their improvement over the unit-norm LS and KL methods was around 2 to 4 dB. As expected, we do not observe any differences between the unit-norm LS and KL methods although the first used eigenvalue decomposition while the second employed SVD realization. The WLS methods were also superior to the STLS scheme because the latter could attain optimum performance only for very high SNRs. Furthermore, the threshold performance of the monic WLS scheme was better than that of the unit-norm WLS method which may be due to the numerical instability in performing the generalized eigenvalue decomposition at larger noise conditions.

The corresponding biases, which were obtained by subtracting the true values from the mean estimates, are plotted in Figures 4 to 6. It is seen that all methods except the STLS scheme were

approximately unbiased when the output noise level was less than 0 dB. The results suggest that the STLS method is only suitable for use in very small noise conditions.

Figures 7 to 12 plot the MSEs and mean errors for a_1 , a_2 and b_0 , respectively, of the unit-norm LS, unit-norm WLS, monic WLS, OE, PE and STLS methods when the system consisted of output noise only at $N = 50$. It is observed that the unit-norm WLS and monic WLS schemes were comparable with the two realizable optimum benchmarks, namely, the OE and PE methods, particularly when the noise level was less than 0 dB. As expected, the unit-norm LS approach performed the worst among all other algorithms. We also observe that the two WLS methods were more robust than the STLS scheme at larger noise conditions, although the monic scheme again had a better threshold performance.

The above two tests were repeated for a larger sample size of $N = 500$ and the results are shown in Figures 13 to 24. Again, the optimality of the unit-norm WLS and monic WLS estimators, in terms of mean and mean square error performance, was illustrated for sufficiently small noise conditions. Furthermore, we see that the MSEs of the WLS methods could reach the CRLB for larger noise levels when N was increased from 50 to 500.

4 Concluding Remarks

The problem of unbiased system identification in the presence of white Gaussian input and output noise, assuming that the noise power ratio is known, has been addressed. The relationship between the unit-norm constraint approach and the Koopmans-Levin method for impulse response estimation has been illustrated via the development of a unit-norm least squares algorithm. This algorithm is then improved by using the technique of weighted least squares (WLS), and the resultant parameter estimates are obtained via the minimization of a WLS cost function subject to a generalized unit-norm constraint. The weighting matrix is the Markov estimate. We also derive another WLS method for system identification based on the monic constraint, which is in fact a relaxation algorithm for solving the maximum likelihood estimate. Computer simulations show that the estimation performance of the two proposed algorithms is very similar and attains optimum performance in terms of mean square errors for both noisy input-output and output-noise-only systems at sufficiently large signal-to-noise ratio condition. Apparently, a challenging extension of this work is to study the convergence issues of the iterative WLS algorithms.

In case of unknown noise power ratio, suitable candidates for the problem are joint output and Frisch methods. The former can achieve maximum likelihood estimation performance but involves extensive computations while the latter is relatively computationally attractive with suboptimum performance [7]. For known ratio of white input and output noise powers, the monic and unit-norm WLS methods outperform Koopmans-Levin (KL) and structured total least squares (STLS)

approaches in terms of mean square error and threshold SNR performance, respectively, although the KL technique is the simplest while the computational complexities of proposed and STLS methods are comparable. When there is either input or output noise, the proposed methods will be the proper choices over the prediction error and output error approaches only when the data length is sufficiently small because the computational load of the latter does not increase significantly with the number of received samples while the WLS solutions do.

References

- [1] S.Haykin, *Adaptive Filter Theory*, Upper Saddle River, NJ: Prentice-Hall, 2002
- [2] L.Ljung, *System Identification - Theory for the User*, Englewood Cliffs, NJ: Prentice-Hall, 1987
- [3] T.Soderstrom and P.Stoica, *System Identification*, Englewood Cliffs, NJ: Prentice-Hall, 1989
- [4] R.Pintelon and J.Schoukens, *System Identification: A Frequency Domain Approach*, IEEE, 2001
- [5] T.Soderstrom, "Identification of stochastic linear systems in presence of input noise," *Automatica*, vol.17, pp.713-725, 1981
- [6] S.Beghelli, R.Guidorzi and U.Soverini, "The Frisch scheme in dynamic system identification," *Automatica*, vol.26, pp.171-176, 1990
- [7] T.Soderstrom, "Accuracy analysis of the Frisch estimates for identifying errors-in-variables system", *Proc. IEEE Conference on Decision and Control*, pp.4281-4286, Dec. 2005, Seville, Spain.
- [8] B.De Moor, "Structured total least squares and L_2 approximation," *Lin. Alg. and its Appl.*, vol.188-189, pp.163-205, 1993
- [9] P.Lemmerling, N.Mastronardi and S.Van Huffel, "Fast algorithm for solving the Hankel/Toeplitz structured total least squares problem," *Numerical Algorithms*, vol.23, pp.371-392, 2000
- [10] S.Van Huffel and J.Vandewalle, "Analysis and properties of the generalized total least squares problem $Ax \approx B$ when some or all columns in A are subject to error," *SIAM J. Matrix Anal.*, vol.10, no.3, pp.294-315, 1989
- [11] T.J.Koopmans, *Linear Regression Analysis of Economic Time Series*, N.V.Haarlem, The Netherlands: De Erven F. Bohn, 1937
- [12] M.J.Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. Automatic Contr.*, vol.9, pp.229-235, July 1964

- [13] K.V.Fernando and H.Nicholson, "Identification of linear systems with input and output noise: the Koopmans-Levin method," *IEE Proc. Pt.D*, vol.132, no.1, pp.30-36, Jan. 1985
- [14] W.X.Zheng, "Noisy input-output system identification - the Koopmans-Levin method revisited," *Proc. 41st IEEE Conf. Decision and Control*, pp.636-637, Las Vegas, Nevada USA, Dec. 2002
- [15] P.A.Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, New York : M. Dekker, 1995
- [16] R.A.Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Processing*, vol.42, no.6, pp.1397-1412, June 1994
- [17] B.E.Dunne and G.A.Williamson, "QR-based TLS and mixed LS-TLS algorithms with applications to adaptive IIR filtering," *IEEE Trans. Signal Processing*, vol.51, no.2, pp.386-394, Feb. 2003
- [18] B.E.Dunne and G.A.Williamson, "Analysis of gradient algorithms for TLS-based adaptive IIR filters," *to appear in IEEE Trans. Signal Processing*
- [19] M.Aoki and P.C.Yue, "On a priori error estimates of some identification methods," *IEEE Trans. Automatic Contr.*, vol.15, no.5, pp.541-548, Oct. 1970
- [20] E.Karlsson, T.Soderstrom and P.Stoica, "The Cramer-Rao lower bound for noisy input-output systems," *Signal Processing*, vol.80, pp.2421-2447, 2000
- [21] G.C.Goodwin and R.L.Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, New York: Academic Press, 1977
- [22] H.N.Kim and W.J.Song, "Unbiased equation-error adaptive IIR filtering based on monic normalization," *IEEE Signal Processing Letters*, vol.6, no.2, pp.35-37, Feb. 1999
- [23] R.Lopez-Valcarce and S.Dasgupta, "A new proof for the stability of equation-error models," *IEEE Signal Processing Letters*, vol.6, no.6, pp.148-150, June 1999
- [24] L.Ljung, *MATLAB System Identification Toolbox, Version 4*, The Math Works, Inc., 1995

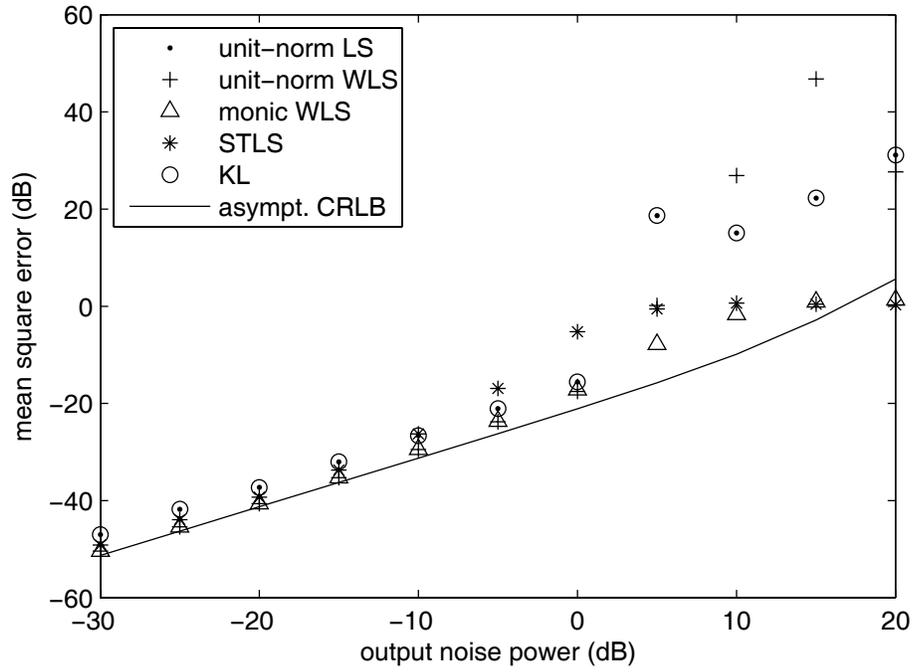


Figure 1: Mean square errors of a_1 for noisy input-output system at $N = 50$

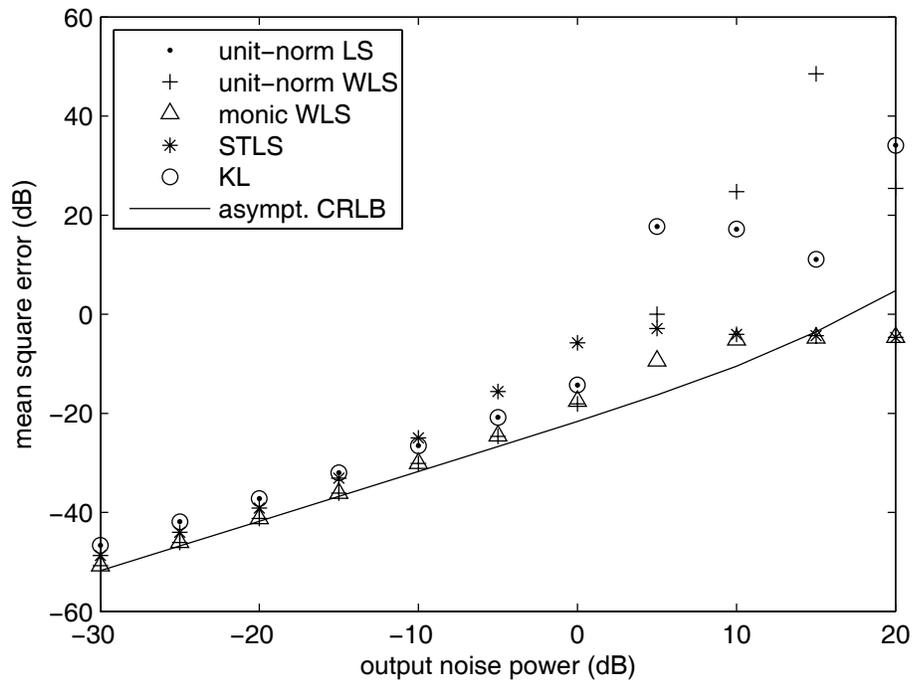


Figure 2: Mean square errors of a_2 for noisy input-output system at $N = 50$

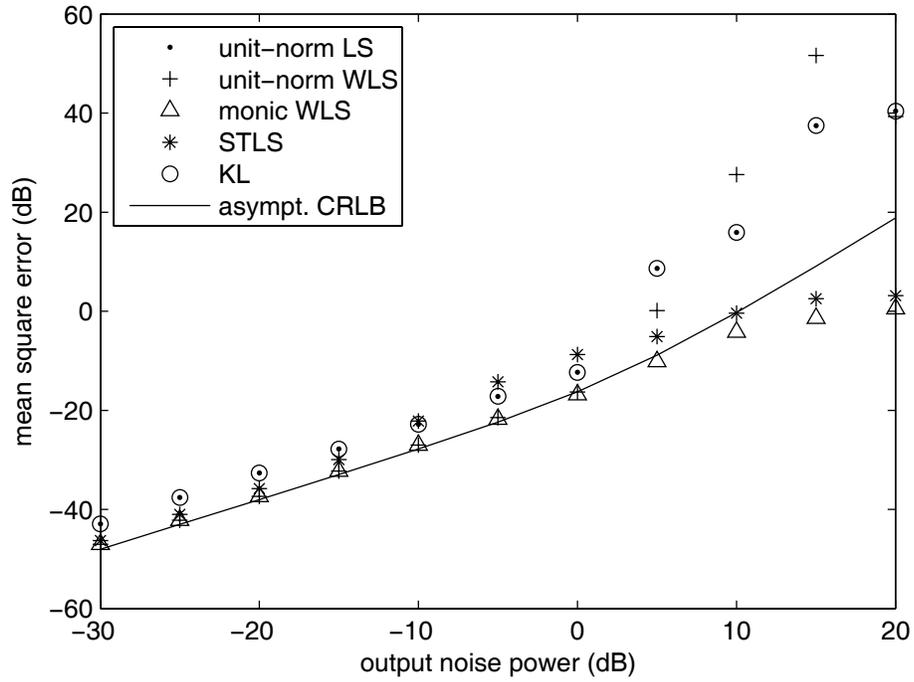


Figure 3: Mean square errors of b_0 for noisy input-output system at $N = 50$

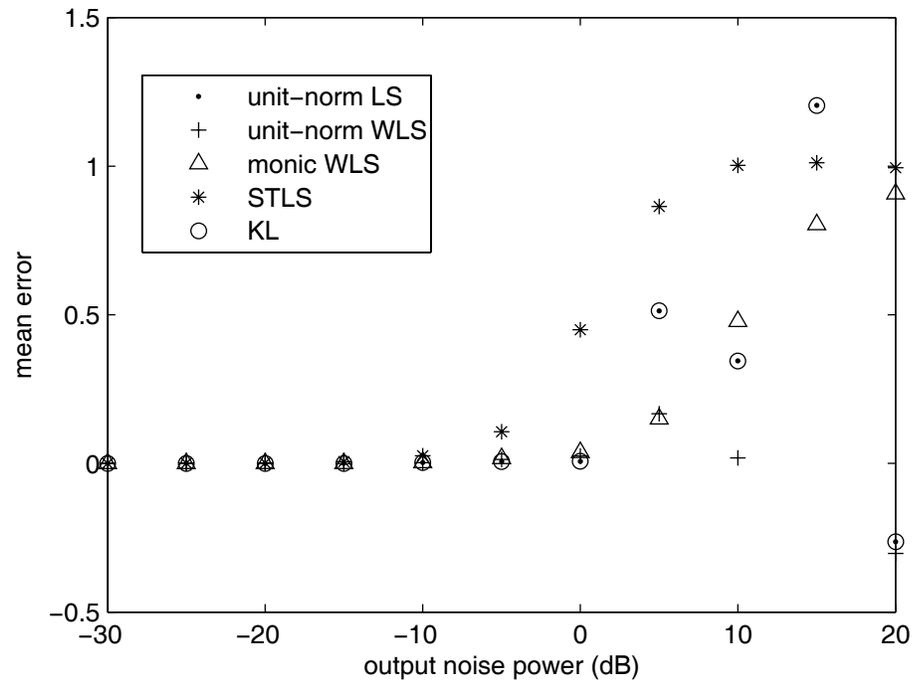


Figure 4: Mean errors of a_1 for noisy input-output system at $N = 50$

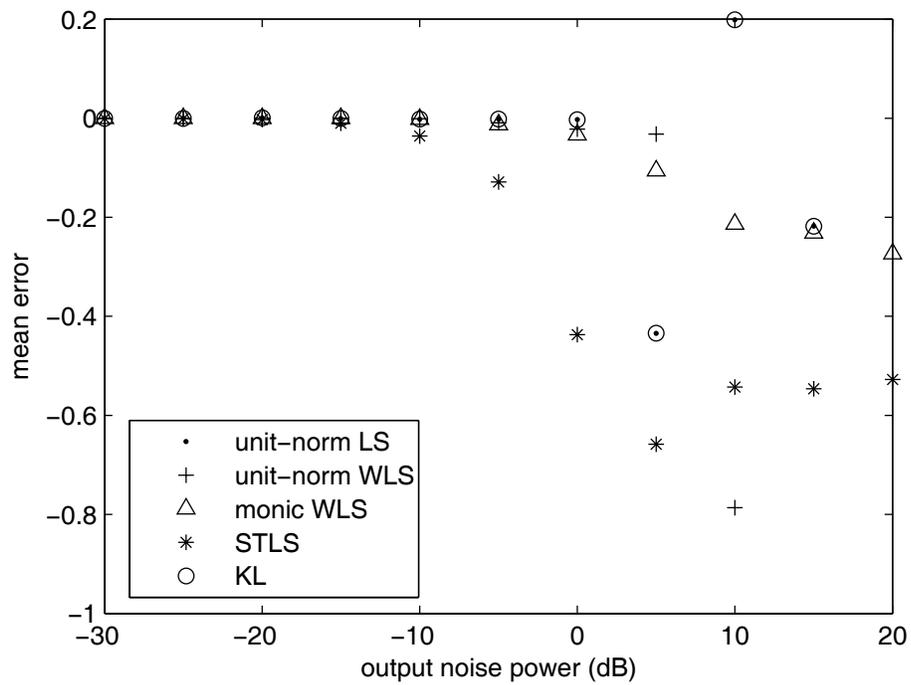


Figure 5: Mean errors of a_2 for noisy input-output system at $N = 50$

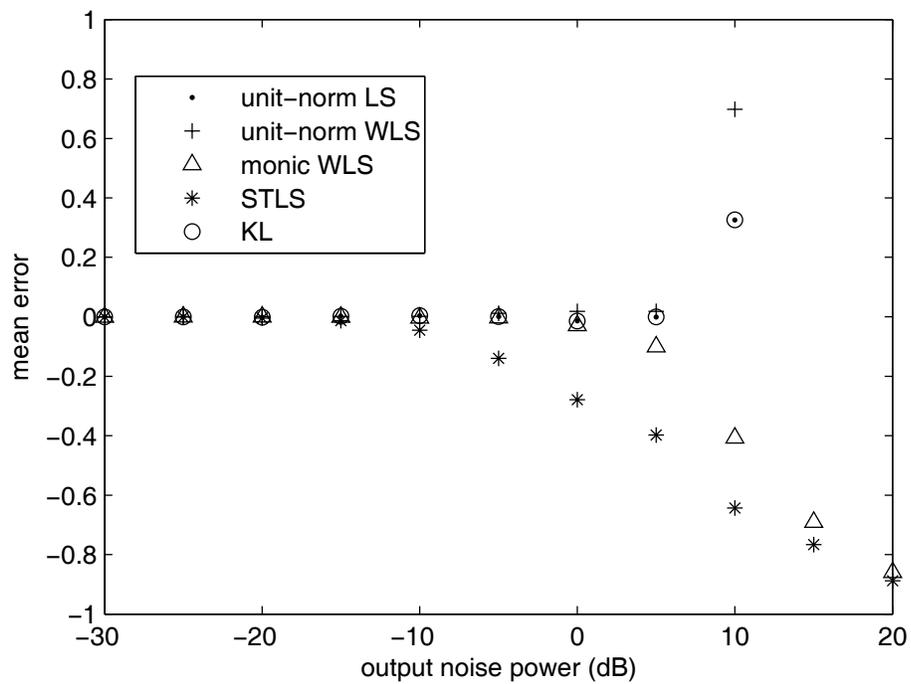


Figure 6: Mean errors of b_0 for noisy input-output system at $N = 50$

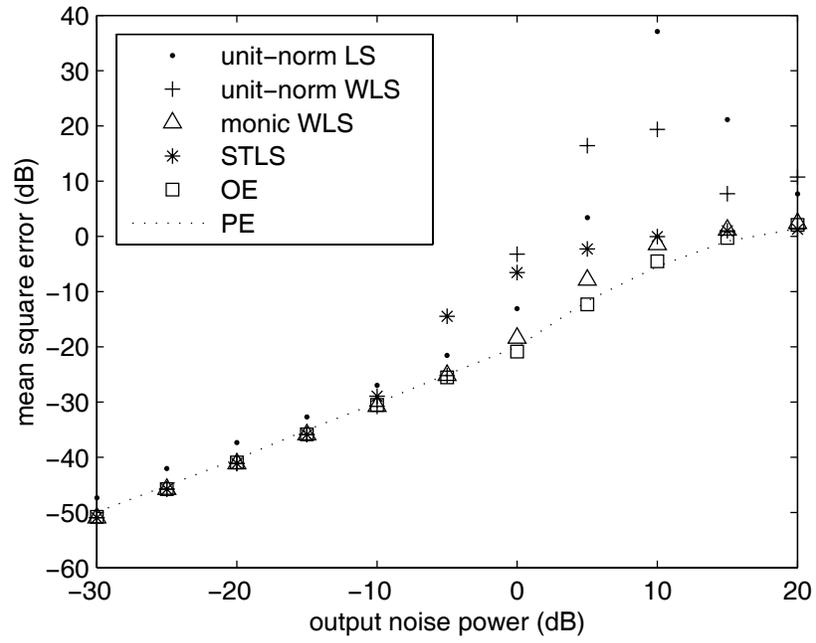


Figure 7: Mean square errors of a_1 for output-noise-only system at $N = 50$

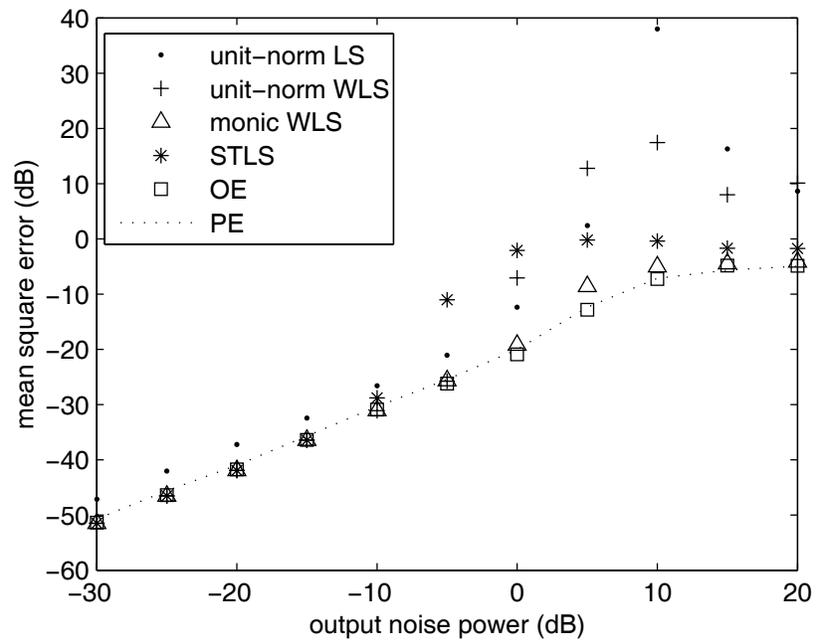


Figure 8: Mean square errors of a_2 for output-noise-only system at $N = 50$

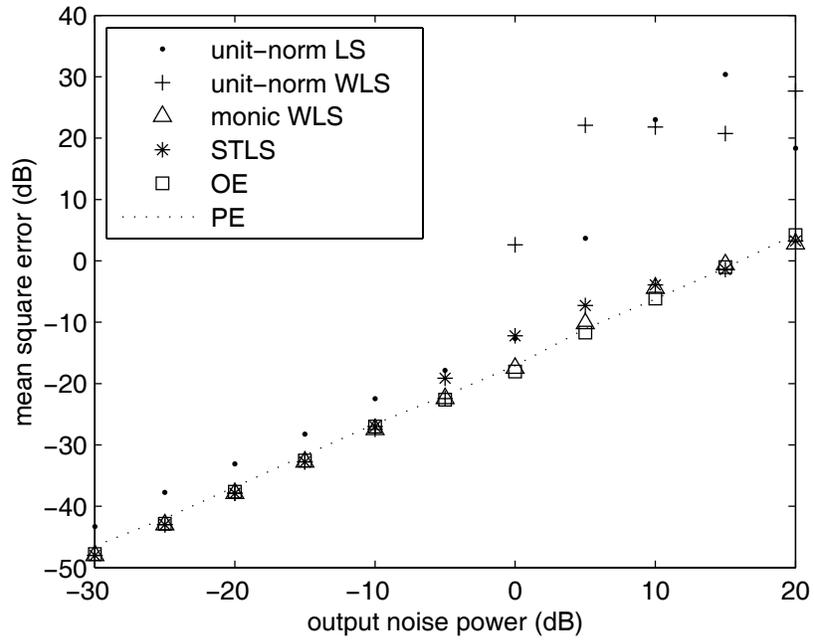


Figure 9: Mean square errors of b_0 for output-noise-only system at $N = 50$

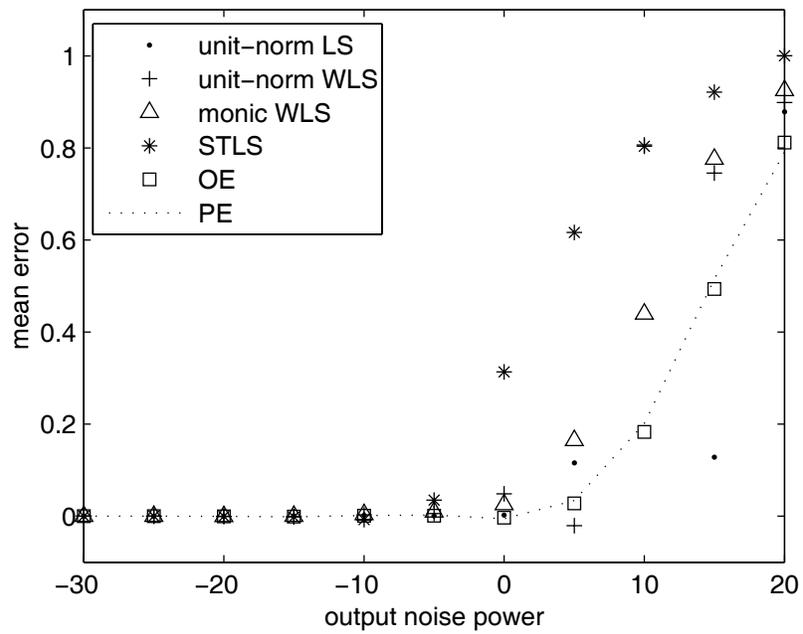


Figure 10: Mean errors of a_1 for output-noise-only system at $N = 50$

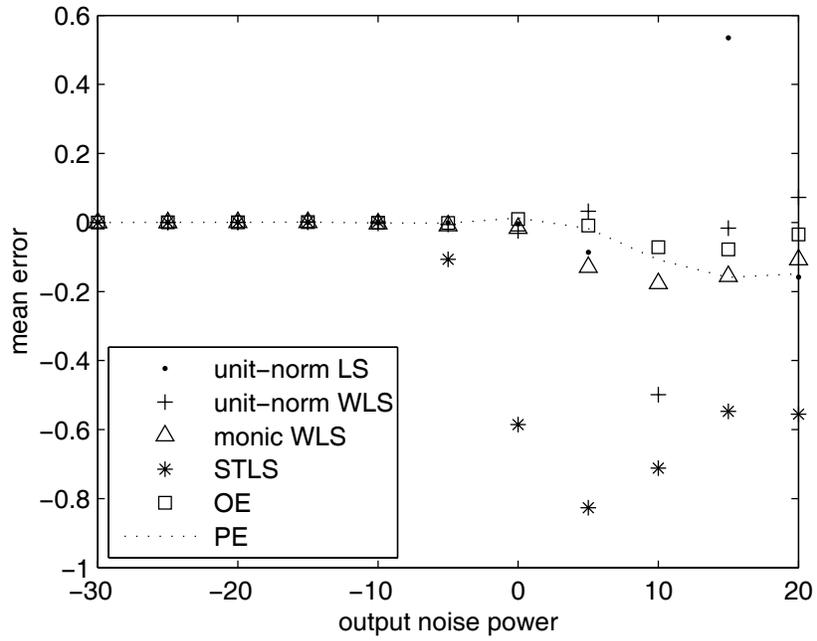


Figure 11: Mean errors of a_2 for output-noise-only system at $N = 50$

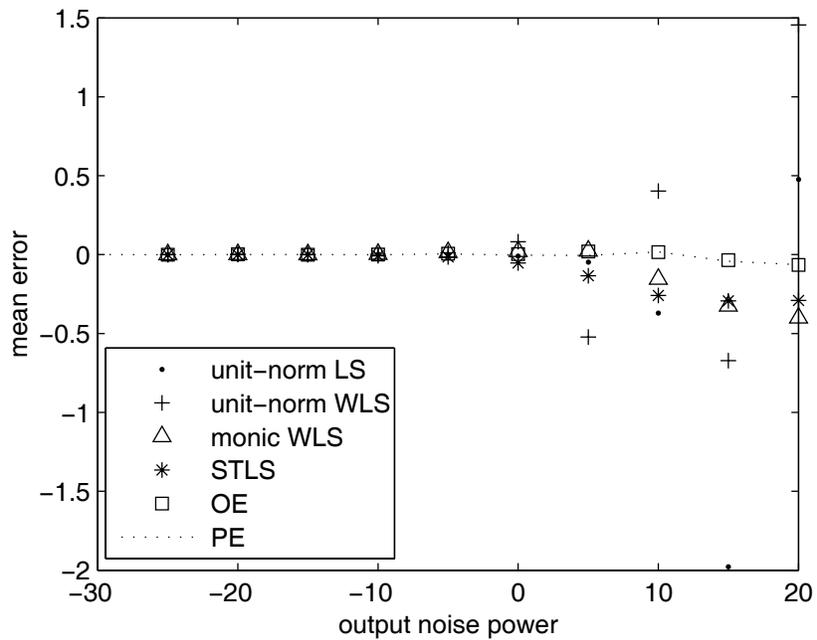


Figure 12: Mean errors of b_0 for output-noise-only system at $N = 50$

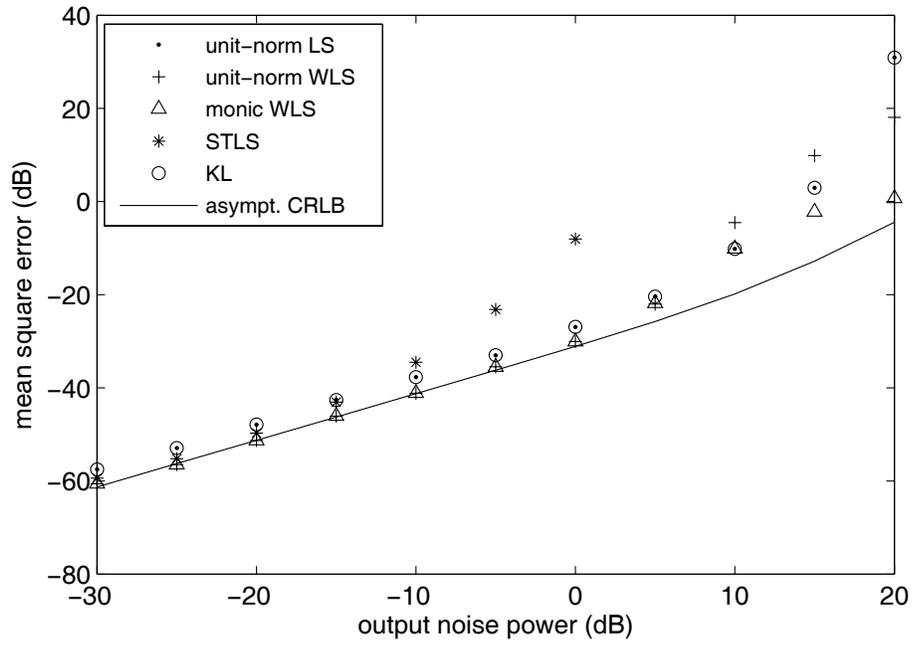


Figure 13: Mean square errors of a_1 for noisy input-output system at $N = 500$

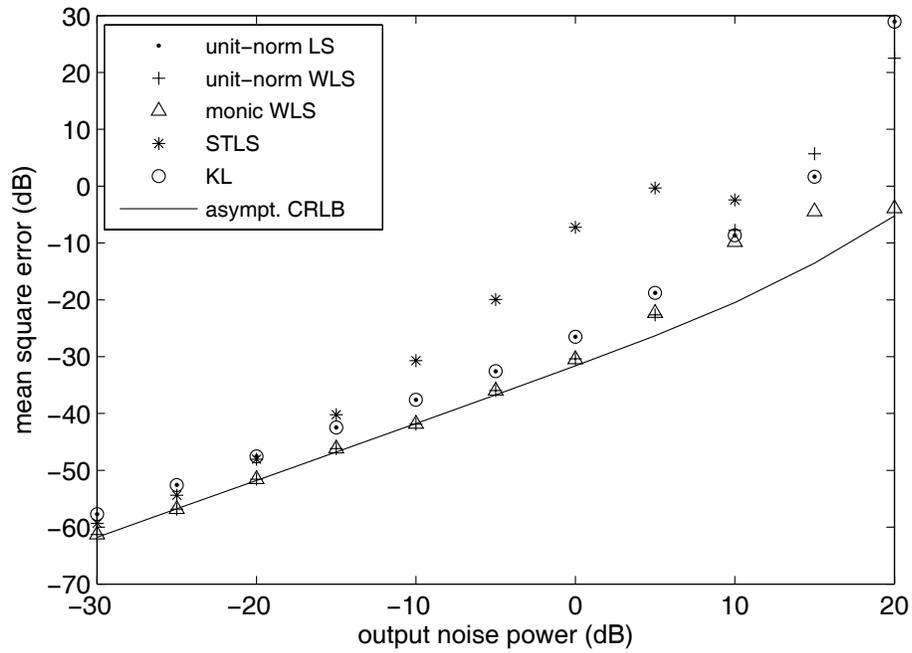


Figure 14: Mean square errors of a_2 for noisy input-output system at $N = 500$

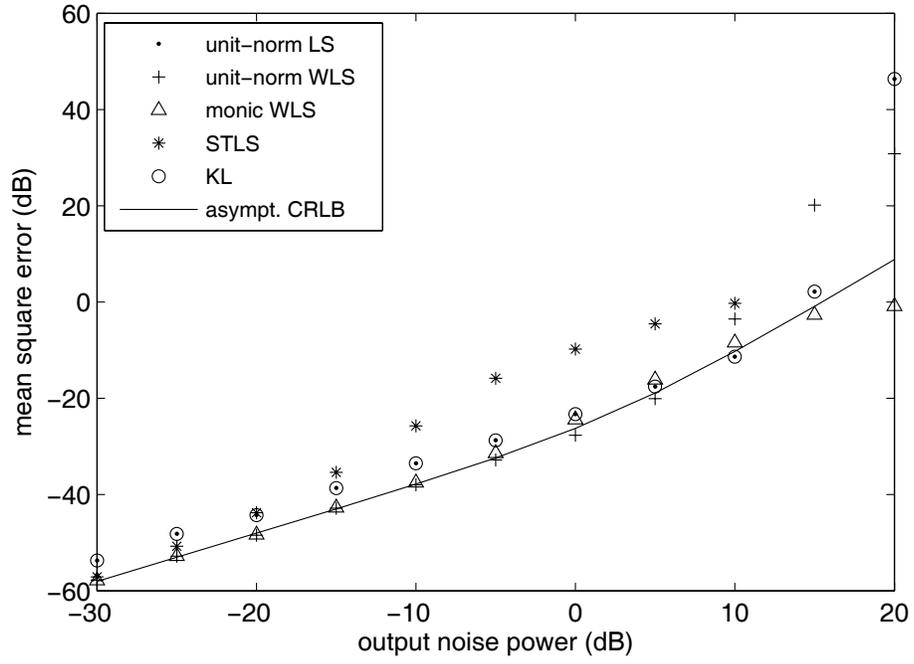


Figure 15: Mean square errors of b_0 for noisy input-output system at $N = 500$

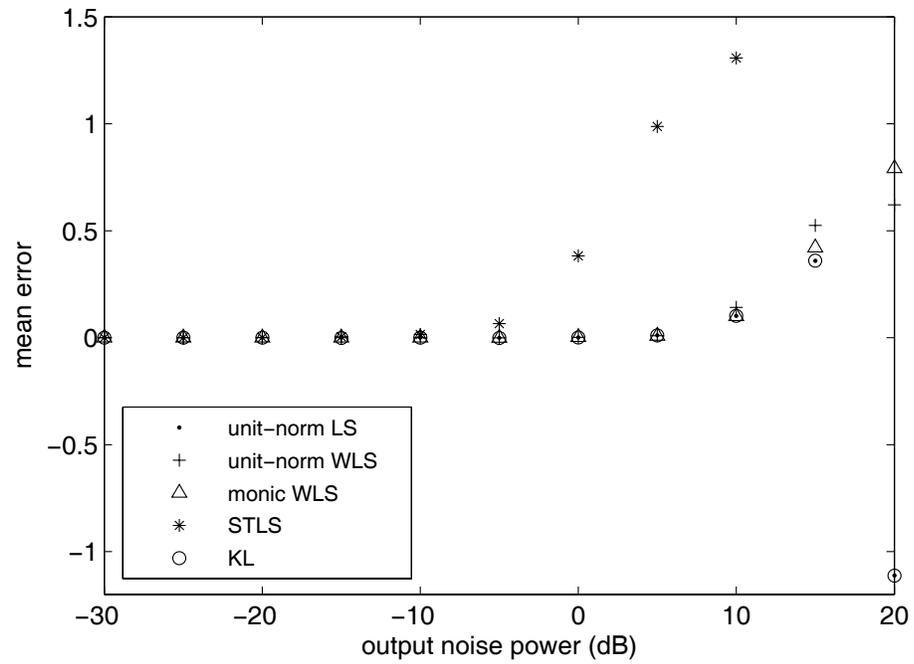


Figure 16: Mean errors of a_1 for noisy input-output system at $N = 500$

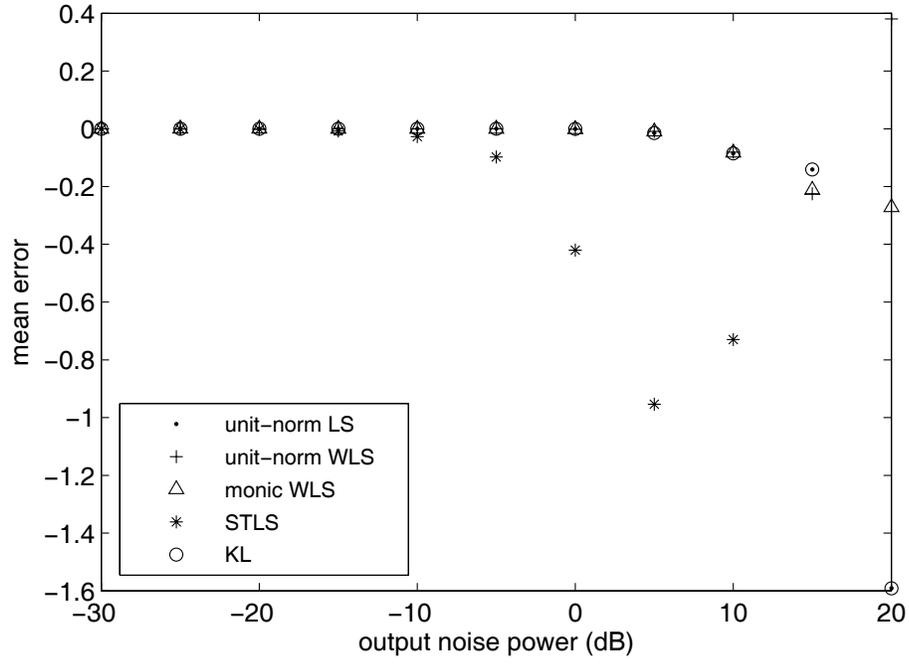


Figure 17: Mean errors of a_2 for noisy input-output system at $N = 500$

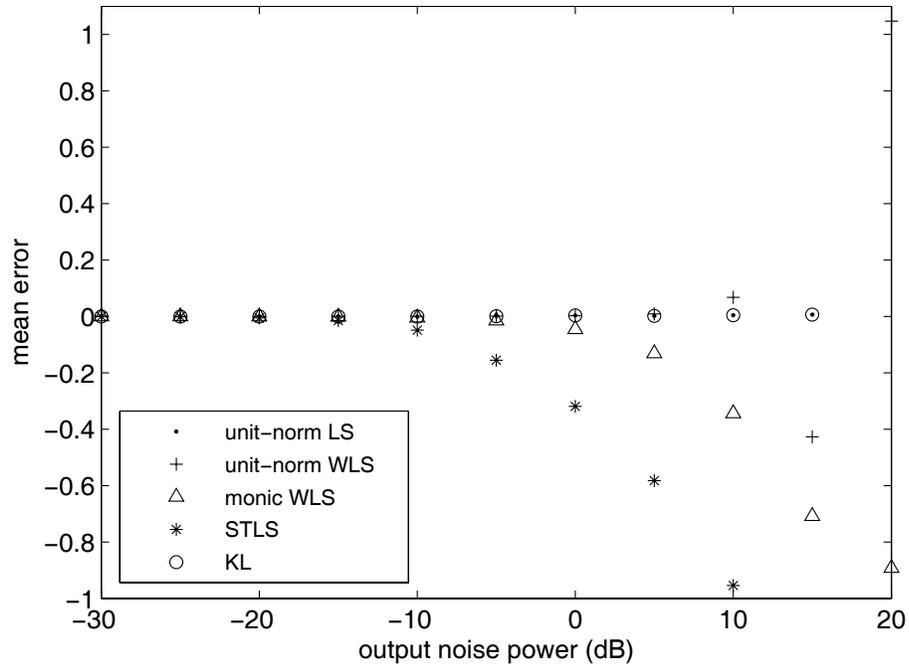


Figure 18: Mean errors of b_0 for noisy input-output system at $N = 500$

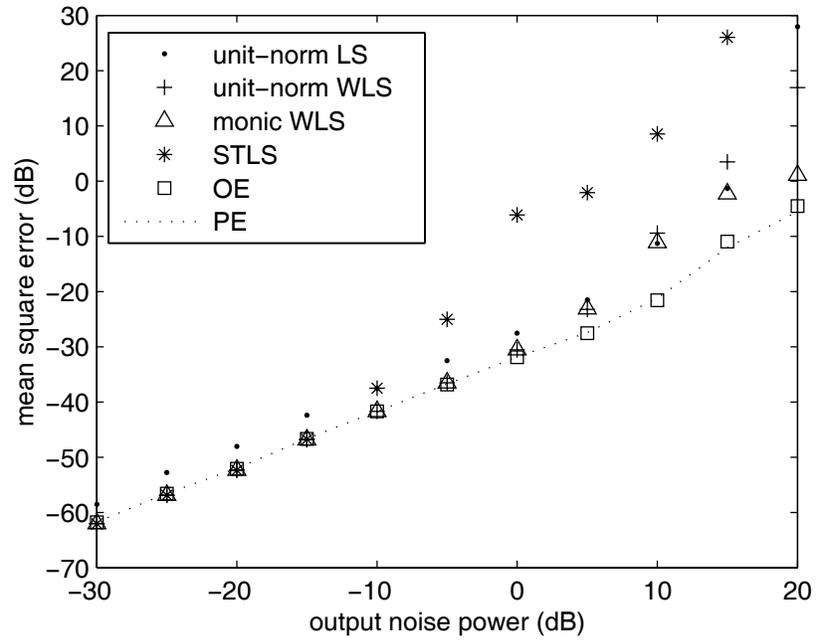


Figure 19: Mean square errors of a_1 for output-noise-only system at $N = 500$

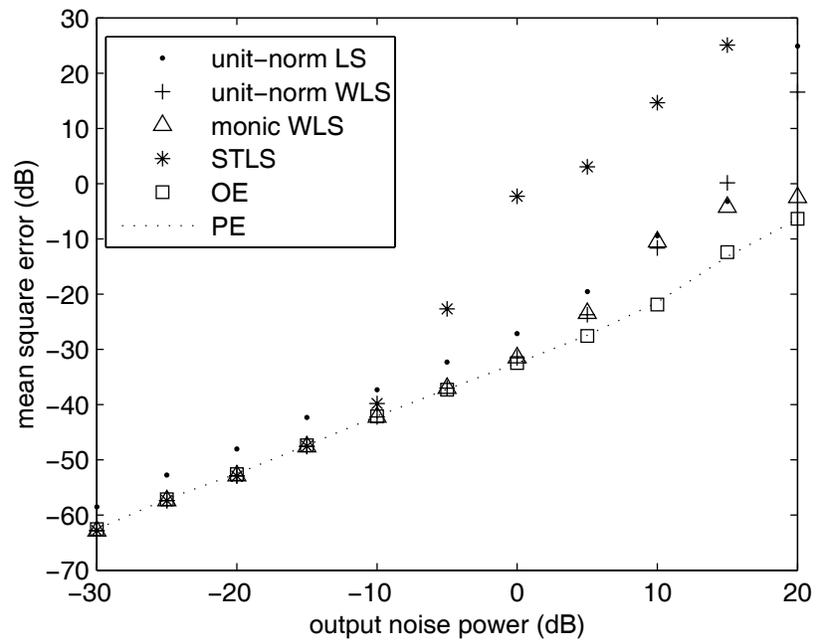


Figure 20: Mean square errors of a_2 for output-noise-only system at $N = 500$

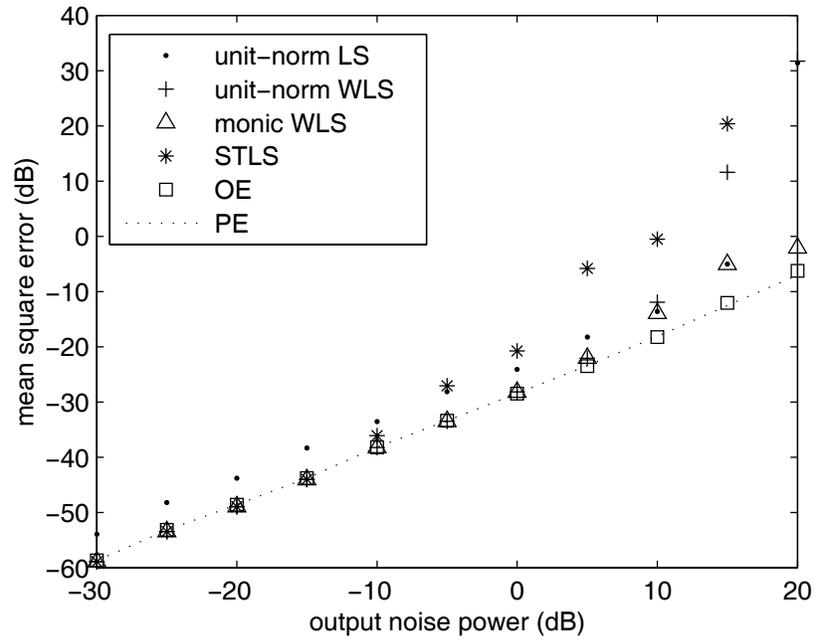


Figure 21: Mean square errors of b_0 for output-noise-only system at $N = 500$

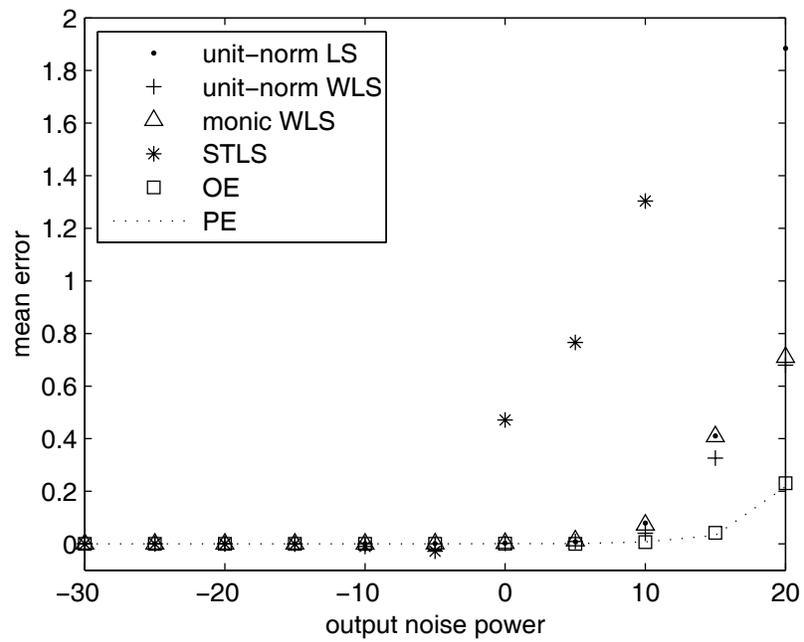


Figure 22: Mean errors of a_1 for output-noise-only system at $N = 500$

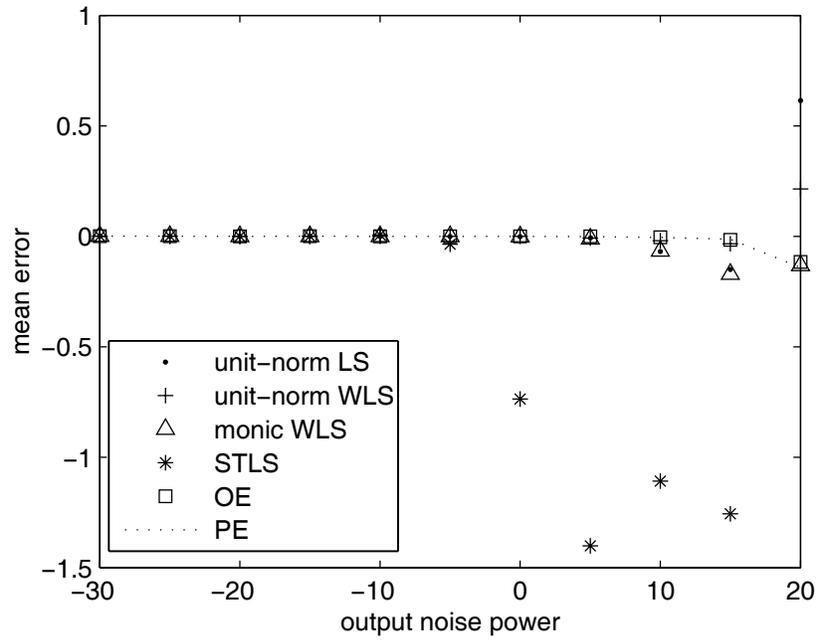


Figure 23: Mean errors of a_2 for output-noise-only system at $N = 500$

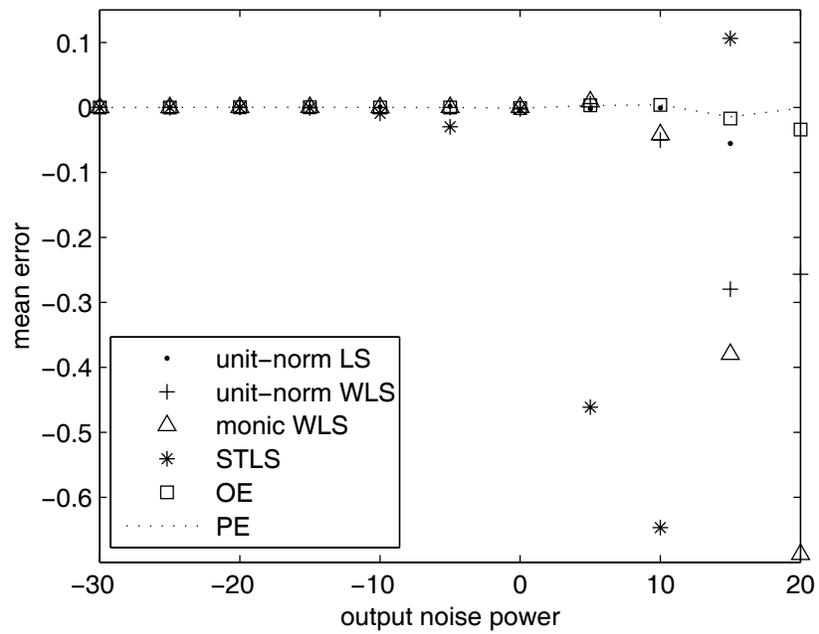


Figure 24: Mean errors of b_0 for output-noise-only system at $N = 500$